

Managing Art History Fuzzy Dates: An Application in Historico-Geographical Authority

*Oreste Signore, R. Barioli**

Art history data management leads to express dates in a »fuzzy« way, which is not suitable for electronic processing. A standardized easy to use way of coding dates, maintaining all the semantics of the »natural« format, has been defined. Each date is, therefore, expressed (in the extended way) as a string of characters, with a well defined syntax, while it is really managed as a set of fields, which contain in a coded way, the same information that are contained in the extended format. The coded format of dates is easily and transparently managed by every Data Base Management System, allowing selection by exact values or by ranges. This conventional way of managing dates has been tested in the implementation of a DBMS based historico-geographical authority file.

The Traditional Approach

Very often, in managing art history data, the user is faced with dates expressed in a fuzzy way (i.e. century, before, after, circa, etc.). while sometimes it is required to find the information referring to some specific year, or range of dates. In addition, it may happen that the date itself is known with a certain approximation (i.e. certainly after 338 B.C. and before 314 B.C.).

However, even if dates expressed in this traditional way are meaningful to the scholars and easy to specify, nevertheless they present some disadvantages. In fact, they are language dependent and hardly managed by computer programs. As a matter of fact, when dates are expressed in the traditional way, it is impossible to join different information on the basis of the time span they apply.

* Address all communications to Oreste Signore and R. Bartoli, CNUCE Insitut of CNR, Via S. Maria 36, I-56100 Pisa, Italy.

The extended format syntax

The main goal in defining a syntax for dates has been to preserve the semantic richness of the traditional fuzzy format, retaining the maximum degree of flexibility, without introducing any useless or unnatural formalism.

The decision has been made (1) to express every date in three possible ways:

- as a simple date;
- as an interval: (Datemin - Datemax)
- as an interval, with indication of a probable date: Dateprob (Datemin - Datemax)

The first form is the correct one of a date known with certainty. The second is used when the scholar is not able to determine a probable date between the minimum and maximum date. The third case comes up when, together with a probable date, there is present conflicting information, but from reliable source.

The preferable forms are, obviously, the first and the third. There are, however, numerous cases where it is not possible to suggest a probable date in that any hypothesis would have little substantiation.

Similarly, there exist cases in which it is not scientifically correct to guess at limits for a date qualified as »circa« (about).

Consequently, after several attempts at formalization, it has been considered opportune to use a series of conventions to represent, in the most natural way for the scholar, the possible forms to give to dates. (2)

On the basis of this convention, every date is expressed in one of the forms recorded above, where Dateprob, Datemin e Datemax all have the same format, which can be one of the following:

- a number, preceded by the expression anno (if the date is known with certainty) or the expression circa (if it is approximate);
- a roman numeral preceded by the expression sec. or inizio or meta' or fine.

They will be followed by the expression a.C. or the expression d.C. (or nothing) depending if the date is before or after Christ.

In the case in which the date is merely hypothetical and is recorded only for the purpose of avoiding inconsistent answers, the expression is followed by »(?)«.

The expression oggi indicates the present year.

The Coded Format

In a database approach, processing and sorting of dates can easily be done if dates are stored as signed numbers: the dates before Christ may be recorded as negative numbers.

As each date is expressed (in the extended way) as a string of characters with a well defined syntax, the basic problem has been to translate every date in a set of fields which contains, in a coded way, the same information that are contained in the extended format.

Each date is really managed as a set of fields which contains, in a coded way, the same information that are contained in the extended format.

In the data entry phase, the operator enters the dates in the extended format.

After loading the database, the whole database is processed in order to update the fields that contain the coded form of the dates.

At the query stage, all the processing is performed using only the coded format, while the results are showed using the extended format.

This approach has been adopted only for speeding up the first pilot phase, in order to check the effectiveness of the adopted solutions. We are working to the routine which will translate the coded form into the extended format, which will also allow different languages to be used.

As soon as the routine is working, the extended format column will be deleted from the database, so allowing a considerable saving in space and a 3NF design of the database.

To be clearer, each date is split into three triplets:

1. Dp type Dp QM Dp num
2. Dmin type Dmin QM Dmax num
3. Dmax type Dmax QM Dmin num

which represent, respectively, the probable date, the minimum date and the maximum date.

The different fields have the following meaning:

- Dx type

indicates if the date is known as certain, is approximate, or is simply a century (eventually a century fraction) and may equal:

Y when the date is known as certain (e.g. certainly on 1820)

A when the date is known as approximate (e.g. circa 1530)

S when the date is given as a century or a fraction of a century (e.g. secolo XVI)

C when the date is computed on the basis of the supplied information (e.g. the minimum and maximum for an exact date, or the probable date for a date given as an interval)

- D x Q M
indicates if the date has been got from a source or is simply an hypothetical date. In this last case, the field value equals »?«
 - Dx num
is the numeric equivalent of the date, and is simply a binary number, which is less than zero for B.C. dates. When a date is expressed as a century or a fraction of century and, therefore, indicates a range (e.g. SECOLO IV means the period from 300 to 399), the value is put to the minimum or the maximum of the range, depending on whether the date is a from date or a to date (e.g. INIZIO III A.C. is put to -299 if it is a from date, and to -276 if it is a to date).
- In the following table, a sample of the equivalence between extended format and coded format is reported. (3)

Date (extended format)	DP	DMIN	DMAX
sec. VIII a.C.	S -750	S -799	S -700
(sec. VIII a.C. - sec. VII a.C.)	C? -699	S -799	S -600
(meta' VII a.C. - meta' VI a.C.)	C? -599	S -654	S -544
sec. VI a.C.	S -550	S -599	S -500
(sec. VI a.C. - sec. V a.C.)	C? -499	S -599	S -400
inizio V a.C.	S -494	S -499	S -489
sec. V a.C.	S -450	S -499	S -400
(sec. V a.C. - sec. IV a.C.)	C? -399	S -499	S -300
(anno 497 a.C. - anno 492 a.C.)	C? -494	Y -497	Y -492
(anno 495 a.C. - sec. IV a.C.)	C? -397	Y -495	S -300
anno 492 a.C.	Y -492	Y -492	Y -492
fine IV a.C.	S -304	S -309	S -300
(circa 211 a.C. - circa 90 a.C.)	C? -150	A -211	A -90
(sec. VII (?) - sec. VIII (?))	C? 699	S? 600	S? 799
(sec. VII - sec. VIII)	C? 699	S 600	S 799

An application: historico-geographical authority

In short, the problem of defining an historico-geographical authority, may be illustrated by the following example.

- The locality presently known as Sezze was belonging to:
 - Stato della Chiesa from 1 do not know when between the Xth and the XIth century up to year 1404
 - Regno di Napoli from year 1404 to year 1414
 - Stato della Chiesa from year 1414 to year 1870
- The locality presently known as Sezze was named:

- Setia from year 382 b.C. to the XIth century
- Castrum Sitiense from the XIIIth century to some unknown year, but probably up to a period that may be estimated to lie between the XVIth century and the XVII century
- Secia from year 1478 to (I may guess) XVIIth century
- How was it named when it was belonging to **Regno di Napoli**.'
- May I have a map showing the evolution of Regno di Napoli with localities with the historical name? (i.e. May I produce an historical atlas?)

It should be noted that the two pieces of information (administrative belonging and historical name) are completely independent and may be got from different sources in different times.

A liaison between them may only be established on the basis of the time.

As a test bed for the proposed formalism for the representation of dates, an historico-geographical authority file has been implemented (4) where, for each locality, information about historical names, historico-political and ecclesiastic structures, the locality was belonging to, are separately recorded.

At present, only the region Lazio, with province of Latina in full (31 Comuni and 269 Localita'), other provinces partially, is covered. (5)

In total, 738 localities are covered, with 2676 notices and 3543 references to sources. For each information, it is recorded the period (from-to), the information apply and the sources the information comes from. These information may be easily correlated, so allowing the production of indexes and of historical atlases for any arbitrarily chosen year (for each locality, the present boundaries are available in digitized form, so that maps may be produced).

All data are stored in a Relational Database (using the SQL/DS product) and may be queried online.

Conclusions

As a conclusion, we may notice that the extended format is understandable and easy to use and the coded internal format:

- performs very well;
- allows a very compact representation (only 12 bytes) for every date;
- may be easily extended to century fractions and to more precise dates, where the month or the day are specified (it is only a matter of defining an appropriate wording in the extended format);
- is a convenient way for exchanging data and may be adopted as a standard, overcoming differences in languages.